Self surveying structures

Danny van der Kleij, 2011

Piet Zwart Institute – Networked media programme

Sniff, scrape & crawl

March 25, 2011

Introduction

In a theoretical experiment I will explore the idea of a building or structure that is observing, listening and learning. The observing utilizes the learning part, a process that can evoke decisions. In this way a building becomes a sentient being, it gains the ability to perceive in a beforehand constructed subjective manner.
Assisted by artificial intelligence and multiple methods of self surveillance this sentient building is trying to find its way in indexing and categorizing the behaviour of its human occupants.
I argue that however complex or life like this computer indexing system is, its initial binary system will interfere with human contingency. Using examples from modern methods for machine vision and event discrimination as well as a more philosophical standpoint, I will look at the possibilities and difficulties this building will have.

Observation, from visual surveillance to computer vision

The building starts to observe and listen, for its eyes and ears are cameras, microphones and points of electronic identification usually at the buildings entrance or another form of physical barrier. Whether or not a human is watching along with the building or system itself, these devices are the building's only methods of acquiring information.
The building's eyes and ears are placed by its owners in places usually and preferably unseen_by the building's visitors or occupants, yet the building's visitors and occupants know that these methods of surveillance are present. The standard shape of the panopticon, the dome shaped structure, has lost its necessity with the placement of surveillance cameras. The sentiment of an invisible omniscience can be obtained through the hardly ever seen camera. The mere notion of being watched and surveyed is enough to accomplish the regulating effect of surveillance. In other words, without having knowledge of the entire surveillance system, the visitors will have the feeling that they're being observed, something that proves sufficient in many systems, a very straightforward example being subway stations. Whenever a camera is discovered this feeling of omniscience is not lost, rather reassured. When viewing the building as a being, we should accept that this being is trying to maintain an omniscience within the field of surveillance.

In an Artificial Intelligence journal article on visual surveillance computer scientists Hilary Buxton and Shaogang Gong, stated that:

> Visual surveillance primarily involves the interpretation of image sequences. Advanced visual surveillance goes further and automates the detection of predefined alarm events in a given context. Whilst the definition of where and how an alarm event may occur is required, it is the intelligent dynamic scene and event discrimination which lies at the heart of advanced visual surveillance. (Buxton, Shoagang, 1994, p.431)

The building can never maintain this omniscience through objective recording. The sheer amount of images and data that is fed into the building's storage capabilities, whether digital or sometimes even physical (in the case of magnetic tape) is too much to analyse by others in the short time between the registration of an event occurring within the building and the moment of judgement by others.

To cope with this vast amount of data it needs a form of data analysis. In a way the building gives up its omniscience in the sense of the ubiquitous recording and archiving of itself, and replaces it with a simplistic form of judgement via elements of artificial intelligence(A.I.).

Decision making, feedback and classical artificial intelligence

When a surveillance camera is assisted by artificial intelligence this objectivity of recording and storage of images is lost and replaced by a systematic form of decision making. One could argue that the placement of cameras and the recording of images is never an objective act, however for the building itself it is fundamentally

different when an element of artificial intelligence is introduced. For the owners of the surveillance system it is no longer a simple question of examining the images. The owners of the surveillance system enter a dialogue with the computer. Depending on the way the system is structured beforehand, the judgement of the computer is deemed either more or less important than the images themselves. An example where the judgement of the system is more important then the camera images is when the camera is only activated when a certain threshold of movement is made between two frames. A well known example of such a system would be to detect a rapid movement of cars, in order to enforce traffic laws. This method of triggering a moment of surveillance prevents the users of the surveillance system to see all the other images, that the camera could have taken. A system wherein an event triggers an automated response is a classical form of a cybernetic system. A similar form of an event triggered system used in surveillance of a structure is the Suicide Box developed by the bureau of inverse technology. Although it remains a question whether the Suicide Box system truly existed as they say on their own website 'When the video was exhibited at the 1997 Whitney Biennial of American Art (1997), the curatorial essay identified it as an imaginative work'. Systems similar to the suicide box are at work within our society everywhere. On the bureau's website they explain this system as the following:

> The bit Suicide Box is a motion detection video system designed to capture vertical activity. Unit includes BITcamera, motion capture card, analysis software and utility concealment casing. In standard operation any vertical motion in frame will trigger the camera to record to disk.
>
> Bureau installed the Suicide Box for trial application in range of the Golden Gate Bridge California 1996; an initial deployment period [100 days] metered 17 bridge events. System efficacy: Suicide Box system supplied public, frame-accurate data of a social phenomenon not previously accurately quantified. Box placement was determined to exploit cultural climate and BIT agent proximity; San Francisco is gateway to the Silicon Valley and both Information capital and Suicide capital of the USA. (bit bureau, 1996, http://www.bureauit.org/sbox/)

Although the system was successful to map out a social phenomenon, it is important for our theoretical experiment to see within this system how a decision is made. It clearly states that the standard operation is to activate the recording whenever vertical motion is in frame. The entire system is clueless of the idea of suicide or even the concept of 'human'. This lack of context becomes even more clear when we read the bureau's report on the seagull factor. 'Excessive triggering caused by unanticipated seagull interference initially disabled the Suicide Box in test deployment.' (bit bureau, 1996, http://www.bureauit.org/sbox/#transcript)

Bit bureau, Seagull Trigger, still from BIT video documentation
Suicide Box trial installation,Golden Gate Bridge CA, 1996, 15 min.

The vertical swooping movement of the seagulls near the golden gate bridge caused the camera to be activated.

We can imagine other forms of wrong interpretation of image data from this example, and equally as important we can imagine the Suicide Box system to not activate and not record a suicide, simply because the suicide didn't have any form of vertical motion.
The lack of context is almost as John Searle's Chinese room argument in the real. In Minds, Brains and Programs, John Searle writes out a thought experiment wherein he asks the question can a machine convincingly simulate an intelligent conversation, does it necessarily understand? If we see vertical motion as a formal element, we could ask the same question, does it understand? John Searle argues the following:

> As long as the program is defined in terms of computational operations on purely formally defined elements, what the example suggests is that these by themselves have no interesting connection with understanding. They are certainly not sufficient conditions, and not the slightest reason has been given to suppose that they are necessary conditions or even that they make a significant contribution to understanding. (Searle, 1980, p.421)

It is almost over obvious to state that the Suicide Box doesn't understand anything that has to do with the social phenomenon, namely suicide near the Golden Gate bridge, however in the dialogue between the system and its owners, the system decides whether a recording is made or not.
In this simplistic artificial intelligence setting, the building is not sentient yet, where sentience is defined as consciously perceiving, we are clearly missing a framework for the building to build its own context.

Learning, neural networks and mimicking human thinking

The building has archived and indexed a vast amount of images, and with little human input is constructing a systematic approach to deal with its human occupants. In order for the surveillance system to register and index the social phenomenon of its occupants in a better way, an element of learning is installed. The implementation of such a learning system within the building is done in the light of security efficiency and cost reduction of security personnel, although statistically this is an improvement, it is important to look at the underlying structure of artificial intelligence.
Wherein the previous examples, a total lack of context and the missing understanding thereof was the fundamental problem, in a learning system, this context is artificially created in the form of a database, and learning is mimicked in a digital way in the form of an artificial neural network.
These artificial neural networks are build upon human correction and pattern recognition from an ever growing database. When the building has learned the ability to compare a specific data-set with the occurrence of an event in real time and from there store the new event into its database, we can assume that the building has gained the ability to build its own form of context. The fundamental building blocks of this context are just as far away from the idea of understanding as the classical form of artificial intelligence. More importantly, for the dialogue between surveillance systems and the humans dealing with those systems is that this lack of understanding is also present when we talk about concepts such as nuance and doubt. In making its decisions the system does not know nuance or doubt and it cannot express them either. As Jordan Crandall writes in his latest work The Geospatialization of Calculative Operations, 'Spatial Data Infrastructures (SDIs) that attempt to consolidate and render inter operable all urban phenomena within standardized calculative architecture' (Crandall, 2009, p.2). This standardization leads to an argument of reasoning versus calculation. The system can however not make a decision, and store this moment as an error or anomaly. With this idea Crandall writes on the construction of a surveillance system context:

> Humans are not required to define parameters for the software to recognize behaviour or objects the 'system itself' decides how a human is classified as opposed to a car or animal or any other object. The system, then, observes the scene to learn and identify normal and anomalous behaviours by way of a constant study of the types of objects that exhibit those behaviours in the scene. Learning from experience, it can adapt to changes in the observed environment 'on its own', detecting, tracking and classifying abnormal behaviour that was not previously defined or anticipated activity that might be deemed high-risk or potentially violent. (Crandall, 2009, p.16)

Even though a system similar to what Crandall writes about is able to formulate new abnormal behaviours, it does so by simply matching data sets through a solely statistical approach. It has exactly the same lack of understanding as the Suicide Box example. From a linguistic perspective, it seems almost contradicting to use a statistical approach for an abnormal phenomenon.
One of the strongest counterarguments to the lack of understanding the environment and context within A.I. is a simple does it need to? Daniell Dennett writes the following on 'Cog' a cognitive humanoid robotics project at M.I.T. Artificial Intelligence lab:

> If the day ever comes for Cog to comment to anybody about Chicago, the question of whether Cog is in any position to do so will arise for exactly the same reasons, and be resolvable on the

> same considerations, as the parallel question about the reference of the word "Chicago" in the idiolect of a young child. (Dennett, 1997, p.2)

How feasible this argument is in a situation where a team of human owners can spend time teaching or reprogramming the system, it doesn't work in a setting where a building must handle event discrimination in order to assist human security personnel. I argue that many feelings of technophobia stem from the fear of being standardized into a system where there is no room for contingent behavior.

Conclusion

The building that is trying to perceive and understand its occupants is still far from being sentient although it has almost all the main senses a human has. Yet the building is at a position where it must decide what is happening, it must tell its owners whether and why something is wrong. In simple systems such as fire alarms, the rate of correct decisions is nearly a hundred percent. Yet when these same simple systems deal with human behaviour, all seems simple enough to fall into error. In classical A.I. there is a lack of context, which results in a lack of subtlety, nuance and doubt. Whether its task and system is simple or complex, the building tries to rigidly order human behaviour into abnormal or normal, unlawful or lawful, violent or non-violent.
In order for the systems' errors to be minimized, an element of learning is introduced. However these new systems only solve the building's problems on a time saving level. If the building can learn from its environment, the building's security personnel doesn't need to actively study that same environment any more. The learning system is build from the same elements of A.I. as its more classical predecessors. In every new and abnormal situation, the building remains idle or gives an outcome suited for a previous yet very different situation. Although designed to handle abnormalities it has a hard time dealing with them.
Yet there are reasons why the building is allowed to decide whether a situation needs to be addressed or not, it is precise, reliable and has an overview that none of its human helpers have. Because of the building's high rate of success, it has gained the trust of its makers and owners. Whenever the system detects an anomaly within its environment which it cannot handle yet, there is a tendency to agree with him. Whether this tendency causes paranoia or a feeling of safety, human assistance seems the only way towards a conscious sense of improvement.

References

Searle, J. R. (1980) 'Minds, brains, and programs.' in Behavioral and Brain Sciences 3 (pp. 417-457)

Daniel C Dennett (1997), 'Consciousness in Human and Robot Minds', by Ito,M. Miyashita,Y and Rolls, E.T. (ed) Cognition, Computation, and Consciousness, Oxford University Press.

Crandall, J (2009) 'The Geospatialization of Calculative Operations, Tracking, Sensing and Megacities' in Theory, Culture & Society(SAGE, Los Angeles, London, New Delhi and Singapore)

Buxton, H. and Gong, S. (1995) 'Visual surveillance in a dynamic and uncertain world' in Artificial Intelligence 78 (pp. 431- 459)

Jeremijenko, N and Rich, K (1996) http://www.bureauit.org/sbox,
http://www.bureauit.org/decade/projects.html

Bibliography

Aström, K.J. and Murray, R.M. (2009) Feedback Systems,An Introduction for Scientists and Engineers, Princeton University Press

Chow,T.W.S and Cho, S.Y. 'Industrial neural vision system for underground railway station platform surveillance' in Advanced Engineering Informatics 16 (pp 73-83)

Alison Cawsey (1997), *Neural Networks*, Sections 7.1, 7.2 and 7.6 of *The Essence of Artificial Intelligence*, Prentice Hall.

Marx, G.T. (2005) 'Seeing Hazily (But Not Darkly) through the Lens: Some Recent Empirical Studies of Surveillance Technologies.' in Law & Social Inquiry, Vol. 30, No. 20 (pp. 339-399)